

04259.P072

Patent

UNITED STATES PATENT APPLICATION

for

A METHOD AND APPARATUS FOR
EXTRACTING DIGITAL DATA FROM A MEDIUM

INVENTORS:

Kevin Snow
David Clifford

Prepared by:

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN, LLP
12400 WILSHIRE BOULEVARD
SEVENTH FLOOR
LOS ANGELES, CALIFORNIA 90025
(408) 720-8300

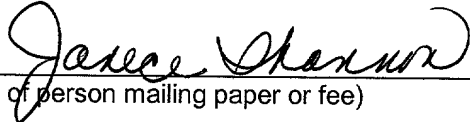
Attorney's Docket No. 004259.P072

"Express Mail" mailing label number EL371009672US

Date of Deposit: November 13, 2001

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231.

Janece Shannon
(typed or printed name of person mailing paper or fee)

 11-13-2001
(Signature of person mailing paper or fee)

A METHOD AND APPARATUS FOR EXTRACTING DIGITAL DATA FROM A MEDIUM

BACKGROUND OF THE INVENTION

Field of the invention

[0001] The invention relates generally to the field of data extraction and encoding, and more particularly to an improved system and method of extracting digital audio data from a medium to one or more playable files.

Background of the Invention

[0002] Digital Audio Extraction ("DAE"), also known generally as "ripping," is the process of copying a track from an audio disc, usually music, to a hard drive or other storage medium by creating a file (or group of files) in any number of encoded and/or compressed formats (e.g., WAV, MP3, . . . etc). A wide variety of software packages that utilize DAE are now available, and the average computer user can easily "rip" any number of tracks from a CD collection to one or more files on a computer hard drive. Subsequently, these tracks can be played back with software designed to read and play extracted audio files.

[0003] Although ripping has become a common practice for many computer users, high quality audio extraction can be difficult because of the complexities inherent in the way data are stored on audio discs. Audio CD data are organized into sectors in order to ensure a constant read rate. Each sector consists of 2,352 bytes of sound data along with synchronization, error correction, and control/display bits. These sectors are further broken down into sound samples. Each sector contains 588 samples of sound

for each of two stereo channels, and each sample contains two bytes (16 bits) of sound data. The standard sampling rate of CD players is 44,100 samples per second.

[0004] Sectors are not arranged in distinct physical units. Instead, the data in one sector are interleaved with data in other sectors so that a defect in the disc will not destroy a single sector beyond correction. In addition, each track's location, or address, is recorded in the disc's Table of Contents ("TOC"), which is stored in the "lead in" area of every disc. Accordingly, an audio disc's TOC, much like a book's, is a good resource for determining where tracks begin and end. The TOC indicates the minute, second, and sector (to $1/75^{\text{th}}$ of a second) at which each track begins.

[0005] Extraction of audio/video content from a compact disk to a hard disk using current DAE software can be a difficult task. Every byte of a 2,352-byte sector of audio data is used strictly for audio. Essentially, no header exists; there is no information in the sector that allows for the exact positioning of a read head over a specific sector. To address an audio sector, a CD-ROM drive uses the TOC data to approximate how far out along the CD it must scan in order to find the beginning of a specified track. Drives typically reach an audio address that is within \pm four sector addresses of the address being sought ($\pm 4/75^{\text{th}}$ of a second in playback time), and a read request may return any one of the nine sectors. This inexact positioning may cause undesired clicks and pops, commonly referred to as "jitter," in extracted audio files.

[0006] Graph 110 of **Figure 1** is a plot (not to scale) of the audio level (e.g., audio volume, audio intensity, audio amplitude . . . etc) of an audio recording 120 over time (horizontal axis). Track divisions 130 represent where tracks (e.g., songs, of audio recording 120) begin and end. Threshold 140 represents a predetermined level

threshold and lines 150 represent the points at which the sound level of audio recording 120 drops above or below threshold 140. For example, the level of audio recording 120 is below threshold 140 during time lapses 152 and 154. Time lapses 152 and 154 represent the dead silence that may exist at the beginnings and ends of songs on a CD, respectively. Lines 160 represent the points at which the sound level of audio recording 120 significantly drops but does not drop below threshold 140. For example, the level of audio recording 120 is dropped significantly during time lapse 164. Time lapse 164 represent a lull in the level of audio recording 120 that may occur between tracks, such as clapping in between songs of a live album. Finally, it is shown that a lull in the level of audio recording 120 does not exist in between tracks 3 and 4. This is an example of two tracks that blend into each other during playback without any lull in sound level.

[0007] Current DAE software can be used to extract audio recording 120, and **Figure 2** shows how current DAE programs function. A current DAE program will extract each track of audio recording 120 separately and create a pulse-code-modulation (PCM) file for each track (PCM files 201-204). These PCM files can eventually be converted to encoded file formats (encoded files 211-214) that may be read for playback of audio recording 120. These encoded file formats may be uncompressed or compressed (e.g., via MP3 or WAV file formats).

[0008] One disadvantage to current extraction techniques is that the software extracts each track from the source CD separately. First the software will read the CD TOC to determine the locations of the tracks to be extracted. Then each track will be extracted from a beginning point that may or may not be where the track actually starts and will end extraction at a point that may or may not be where the track actually ends. Again,

Table 1. The number of cases of acute myocardial infarction by age group and sex in the city of Moscow, 1980-1986

	Male		Female	
Age group, years	No.	%	No.	%
< 35	70	0.1	10	0.02
35-44	1,000	1.5	100	0.2
45-54	2,000	3.0	200	0.4
55-64	3,000	4.5	300	0.6
65-74	4,000	6.0	400	0.8
≥ 75	5,000	7.5	500	1.0
Total	15,000	22.5	1,500	3.0

[illegible]

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] A better understanding of the present invention can be obtained from the following detailed description in conjunction with the following drawings, in which:

[0012] **Figure 1** illustrates a graph of an exemplary audio recording including sound levels plotted over time.

[0013] **Figure 2** illustrates a method by which digital audio extraction is used in the current art.

[0014] **Figure 3** illustrates some of the drawbacks in using current digital audio extraction techniques including some of the causes of jitter.

[0015] **Figure 4** illustrates one embodiment of the digital audio data extraction process, wherein audio data is extracted, segmented, encoded and sliced.

[0016] **Figure 5** illustrates a flow diagram of one embodiment of the digital audio data extraction process.

[0017] **Figure 6** illustrates a flow diagram of another embodiment of the digital audio data extraction process.

[0018] **Figure 7** illustrates a CD/DVD storage and playback system as it is used in conjunction with one embodiment of the invention.

TOCTT 320T660

DETAILED DESCRIPTION OF THE INVENTION

[0019] In the following description, for the purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be apparent, however, to one skilled in the art that the present invention may be practiced without some of these specific details. In other instances, well-known structures and devices are shown in block diagram form to avoid obscuring the underlying principles of the present invention.

[0020] The process of digital audio extraction first requires a source medium that contains digital audio data. The source medium can be one of a variety of plastic disc forms (e.g., CDs, MiniDiscs, DVDs) or magnetic disk forms (e.g., floppy or hard disks). The audio data contained on the source medium is arranged in a format, e.g., pulse code modulation (PCM) format, that can be read by a standard audio player such as CD player. In addition, the audio data is typically divided into tracks, where each track represents a distinct block of audio data such as a song. Source mediums can contain a multitude of tracks and additional data that indicate the start of each track relative to the data in minutes, seconds, and sectors. For example, the Table of Contents ("TOC") of a CD contains data that indicate to a CD player the points at which CD tracks begin.

[0021] **Figure 4** shows graph 110 which is also depicted in **Figures 1 and 2**. Again, graph 110 is a plot (not to scale) of the sound level (vertical axis) of audio recording 120 over time (horizontal axis). Track divisions 130 represent where tracks, e.g., songs, of audio recording 120 begin and end. Threshold 140 represents a predetermined level threshold and lines 150 represent the points at which the sound level of audio recording

120 drops above or below threshold 140. For example, the level of audio recording 120 is below threshold 140 during time lapse 442, 444, and 448. Time lapses 442, 444, and 448 represent the dead silence that may exist at the beginnings and ends of songs on a CD, respectively. Lines 160 represent the points at which the sound level of audio recording 120 significantly drops but does not drop below threshold 140. For example, the level of audio recording 120 is dropped significantly during time lapse 444. Time lapse 444 represents a lull in the level of audio recording 120 in between tracks, such as clapping in between songs of a live album. Finally, it is shown that a lull in the level of audio recording 120 does not exist in between tracks 3 and 4. This is an example of two tracks that blend into each other during playback without any lull in sound level.

[0022] **Figure 5** shows a flow diagram of the steps taken by one embodiment of the invention to extract, analyze, and store digital audio data. It will be appreciated by those skilled in the art that the steps diagramed in both **Figures 5** and **6** may be followed in the indicated order or in a different order while achieving many of the same benefits of the present invention. In addition, certain steps may be skipped and/or alternate steps may be employed while complying with the underlying principles of the invention.

[0023] At block 510 the raw audio data contained on the source medium is read by a computer device such as a CD-ROM, and at block 520 the raw data is stored as raw file 420 in an addressable memory space such as a hard drive or RAM. In one embodiment, rather than storing the raw data as a "file" at 420, the raw data may be streamed from the CD to an application which operates on the data stream as described herein. If a CD is being ripped, this file/stream 420 contains PCM data that includes audio data and possibly additional data (e.g., TOC data) used to address specific

locations within the PCM data. It will be understood that the TOC data read from the source medium can be stored as part of raw file/stream 420 or as a separate file (e.g., TOC file 524).

[0024] Audio recording 120 of **Figure 4** is an exemplary plot of the sound levels produced when the audio data of raw file/stream 420 are played. At block 530, of **Figure 5**, raw file/stream 420 is analyzed to determine the points at which the sound level of audio recording 120 drops below threshold 140. Threshold 140 may be set at a sound level comparable to the threshold of human hearing. In one embodiment, each point analyzed is equivalent to a sound sample (e.g., two bytes of data for each channel), which will represent some amplitude of sound. One embodiment of the invention locates these points using the TOC data stored in raw file/stream 420. The TOC data will indicate at which minute, second, and sector each track of the raw audio data (or other types of multimedia data) begins. For example, track edge 430 includes an address representing where track two begins. Once these locations are known, samples of audio data located at predetermined distances behind and/or ahead of, relative to time, each track edge is analyzed to determine if the indicated sound levels are above or below threshold 140. For example, it will be determined that the data sectors located across time lapse 442 contain sound levels below threshold 140. The data sectors marking the locations where the indicated sound levels cross threshold 140 may be stored in an index file, e.g., index file 535. In addition, if there exists no data sectors that indicate sound levels below threshold 140 in the vicinity of a track edge, then that information is stored as well.

[0025] Once the raw file/stream 420 (or the streamed content) has been completely analyzed, moving onto block 540, one embodiment of the invention divides raw file/stream 420 into smaller chunks or segments of data. This division will occur at every track edge except at those edges where sound levels are maintained above threshold 140. For example, a segment division will occur at track edge 430 but not at track edge 432 or 434. This is because the sound levels represented by time lapse 442 drop below threshold 140 and the sound levels represented by time lapses 444 and 446 are maintained above threshold 140. Sections of audio data that contain sound levels below threshold 140 will be referred to herein as “silent” sections. During or subsequent to the division of raw file/stream 420 (see block 540), the silent sections, represented by time lapses 440, 442 and 448, will be excluded from the resulting segments.

Accordingly, segments 450 and 455 are generated and their corresponding shorter lengths are shown in **Figure 4**. In one embodiment, segments 450 and 455 are merely two smaller files similar or identical in format to raw file/stream 420. In addition, while segment 450 contains only track 1, segment 455 contains tracks 1, 2 and 3. It will be appreciated by one skilled in the art that excluding sections of data that do not benefit the end user, e.g., silent sections, will be advantageous for a variety of reasons. For example, there will be less data to encode, the resulting files will be smaller in size, and the multimedia content stored within the files will start and end more precisely.

[0026] Depending on the number tracks and silent sections contained in raw file/stream 420, a multitude of segments may result after the division at block 540 of **Figure 5**. Again, each of the segments can contain one or more tracks, and if one segment contains more than one track the sound levels produced at the track edges are

high enough to be heard by an end user. At block 550, segments 450 and 455 are encoded to a data format that may be read by a standard digital audio player. For example, segments 450 and 455 may be encoded to a WAV file format or they may be encoded and compressed to an MP3 file format such that a standard MP3 player may read and play them. It will be understood that the encoding step of block 550 may occur before the segmenting step of block 540 or before the analyzing step of block 530 without losing the inventive features described herein. In addition, it will be appreciated that any number of segments produced at block 540 may be encoded and a variety of encoded formats may result following block 550. As shown in **Figure 4**, segment 450 is encoded to playback file 460 and segment 455 is encoded to multi-track playback file 465. It will be understood that the size of these files will be different after encoding and/or compression takes place, but the resulting length of each segment, in time, will be preserved during playback. Either playback file 460 or 465 may be played by a standard digital audio player; however, a user may desire to store one of the three tracks contained on multi-track playback file 465 separately or insert one of the three tracks into a different arrangement of tracks. Therefore, at block 560, multi-track playback files are split such that a separate file is created for each distinct track.

[0027] One embodiment of the invention contains a file slicer that can divide one or more encoded files into their separate tracks. The file slicer can separate two or more tracks from a multi-track playback file such that no pertinent data are lost. This is accomplished with the help of the TOC data found in either TOC file 524 or raw file/stream 420, depending on where the TOC data were stored at block 520. The slicer uses the TOC data to find the precise sector(s) at which each track of playback file 465

begins. Then the slicer divides the playback file into smaller playback files, wherein each smaller playback file will contain a complete and distinct track. The slicer extracts all the corresponding groups of sectors of data that are representative of each track and then stores each extracted group as a smaller playback file. Therefore, no sectors are lost. For example, playback file 465 of **Figure 5** can be sliced into smaller playback files 472, 474 and 476. Each of the smaller playback files contains a complete and distinct track. These smaller playback files may be played separately or in sequence without any noticeable loss of sound quality during the time a player is switching between tracks.

[0028] It will be understood that the track addresses indicated by the TOC data are only compatible with the raw audio data stored in PCM format. Once encoding occurs the format of the audio data will change and the number and arrangement of sectors in the encoded data file may also change. However, one embodiment of the invention will anticipate this change and covert the address given by the TOC data to an address that is compatible with the encoded data format.

[0029] One embodiment of the present invention also includes an audio player. This audio player is capable of reading a variety of encoded file formats including those uncompressed and compressed. The audio player can playback the files generated by block 550 or 560 of the extraction process. One advantage of the audio player is that it can recognize tracks that are generated from a multi-track segment. For example, playback files 472, 474 and 476 are three separate tracks originating from segment 455. These tracks are unique because no sections of silence exist in between any two contiguous tracks. This is common for audio recordings that contain tracks that blend

into each other, without any pauses of silence, or off a live album where a listener can still hear the sounds of the band and audience between songs. Therefore, it is advantageous to playback these tracks in their original order without any undesired silence or jitter. With reference to index file 535, the audio player can determine if two or more tracks are intended to be played end-to-end without a pause of silence, and, if so, the player moves from playing one track to the next without a pause or noticeable loss of sound quality. Furthermore, if a pause of silence is meant to exist between two contiguous tracks (e.g., between tracks 1 and 2 of **Figure 4**), the player can reference index file 535 to determine exactly how long the pause of silence was in the original recording and can reinsert that amount of silence in between the tracks currently being played.

[0030] As seen in **Figure 6**, a variation of the audio data extraction process described above may be executed while achieving the same benefits of the present invention. The steps represented by blocks 610, 620 and 630 are identical to those of blocks 510, 520 and 530, respectively. However, at block 640, the raw audio data is not broken up into segments, but is rather encoded into a large playback file, e.g., playback file 645. It will be understood that playback file 645 may contain all or less than all of the tracks found on raw data file 420. A player, in another embodiment of the invention, may playback any number of tracks off of playback file 645, in any sequence, and can use the information stored in index 535 to determine the exact positions to begin and end playback such that there is no noticeable loss in sound quality.

[0031] In **Figure 7** a CD/DVD storage and playback system 700 is illustrated in another embodiment of the invention. The system 700 includes a storage and playback

TCW/RJA 09/717,458

apparatus 705 which communicates over a network 740 to one or more servers 750.

The steps represented by **Figure 5** or **Figure 6** may be embodied in program code stored on storage device 730 of playback apparatus 705 and may be loaded into memory 715 and executed by a microprocessor 710 contained therein. Playback apparatus 705 may be a stand-alone terminal or a client terminal connected with a network, e.g., network 740, further connected with one or more servers, e.g., server 750, via network interface card 720.

[0032] In one embodiment, a user loads each of his or her CDs into playback apparatus 705 and all of the digital audio data contained on each CD is extracted and stored into separate raw data files on storage device 730. Each raw data file may be segmented, encoded, and sliced given the methods described above. Ultimately, playback files containing individual tracks are generated, and the user may playback such files on the playback apparatus 705. The playback apparatus 705 may also include user display 760 which can display information relevant to a selected track. For example, user display 760 may display the artist name, album title, song title, track number (relative to a track set) and album release date. It will be appreciated that this CD-related information may be retrieved from data stored on the CD, from manual user input, or from a database found on a server 750.

[0033] In one embodiment, the techniques described herein may be implemented on a specialized multi-CD ripper apparatus such as that described in co-pending application entitled "MULTIMEDIA TRANSFER SYSTEM" (serial no. 09/717,458) which is assigned to the assignee of the present application and which is incorporated herein by reference. For example, using this system, content from multiple CDs may be

concurrently processed and stored on a user's playback apparatus 705 in the manner described above.

[0034] One specific embodiment of a system for reading, slicing, encoding and splitting multimedia content as described herein is illustrated in **Figure 8**. The system is illustrated as a plurality of modules which may be embodied in hardware, software, or any combination thereof. As illustrated, the multimedia content 810 is initially streamed from a CD 820 (or other medium) into memory by a media ripper module 830. The sound levels of the raw streamed content are then analyzed by a wave slicer module 840 which separates the raw stream data into segments based on its analysis (e.g., using the techniques described above). Index data, TOC data and/or other types of multimedia-related data may also be processed by the media ripper module 830 and the wave slicer module 840 during the foregoing processing stages. One or more multimedia encoder modules 850 then encode each of the multimedia segments using a particular encoding algorithm (e.g., MP3, AC-3, . . . etc, for audio; MPEG-2, MPEG-4, RealVideo 8, . . . etc, for Video). Once encoded, each individual segment may then be logically divided by a splitter module 860 as described above. For example, if the raw multimedia data is audio data, the splitter module 860 may divide the encoded audio segments into a plurality of distinct MP-3 files. As previously mentioned, audio encoding may occur before or after audio segment splitting, depending on the embodiment.

[0035] Embodiments of the present invention include various steps, which were described above. The steps may be embodied in machine-executable instructions. The instructions can be used to cause a general-purpose or special-purpose processor to

[0036] Elements of the present invention may also be provided as a machine-readable medium for storing the machine-executable instructions. The machine-readable medium may include, but is not limited to, floppy diskettes, optical disks, CD-ROMs, and magneto-optical disks, ROMs, RAMs, EPROMs, EEPROMs, magnet or optical cards, propagation media or other type of media/machine-readable medium suitable for storing electronic instructions. For example, the present invention may be downloaded as a computer program which may be transferred from a remote computer (e.g., a server) to a requesting computer (e.g., a client) by way of data signals embodied in a carrier wave or other propagation medium via a communication link (e.g., a modem or network connection).

[0037] Throughout the foregoing description, for the purposes of explanation, numerous specific details were set forth in order to provide a thorough understanding of the present system. It will be apparent, however, to one skilled in the art, that the system and method may be practiced without some of these specific details. For example, while the techniques described above were employed in the context of ripping audio from CDs, the same techniques may be employed using a variety of different media (e.g., DVD audio/video). Accordingly, the scope and spirit of the invention should be judged in terms of the claims which follow.